Implementing SRM V2.2 functionality in dCache


Timur Perelmutov

for the dCache team

International Conference on Computing in High Energy and Nuclear Physics

Victoria, BC

Project Topology : The Team

dCache

Head of dCache.ORG

Patrick Fuhrmann

Core Team (Desy and Fermi)

Bjoern Boettscher

Alex Kulyavtsev

Iryna Koslova

Dmitri Litvintsev

David Melkumyan

Dirk Pleiter

Martin Radicke

Owen Synge

Vladimir Podstavkov

Head of Development FNAL :

Timur Perelmutov

Head of Development DESY :

Tigran Mkrtchyan

External Development Contribution

Gerd Behrmann, NDGF

Jonathan Schaeffer, IN2P3

Andrew Baranovski, CEDPS

Ted Hesselroth, OSG

Support and Help

Greig Cowan, gridPP

Stijn De Weirdt (Quattor)

Maarten Lithmaath, CERN

Flavia Donno, CERN

Project Topology: Partners

Code contribution

besides DESY, FERMI

NDGF : ftp (protocol V2)

IN2P3 : HoppingManager

Integration. Verification

- CERN
- Open Science Grid
- d-Grid

SRM V1.1 interface

- SRM V1.1 has been a part of dCache for over 4 years
- Used in production by US-CMS for over 2 years
- Solid protocol **but**
- Did not include
  - Explicit Space Reservation and Management
  - Directory functions
  - File Access Permission management
  - Abstractions to describe type and quality of service
- Weak Error and status reporting
- SRM 2 addressed many of the issues

LHC needs

- Common interface to T0, T1 and T2 storage
- Guarantee of space availability
  - Space Reservation
- Storage Class differentiation
  - Access Latency and Retention Policy
- Flexible Namespace management
  - Directory Functions
- ACL Support
  - Permission management functionality
- SRM V2.2 was the answer

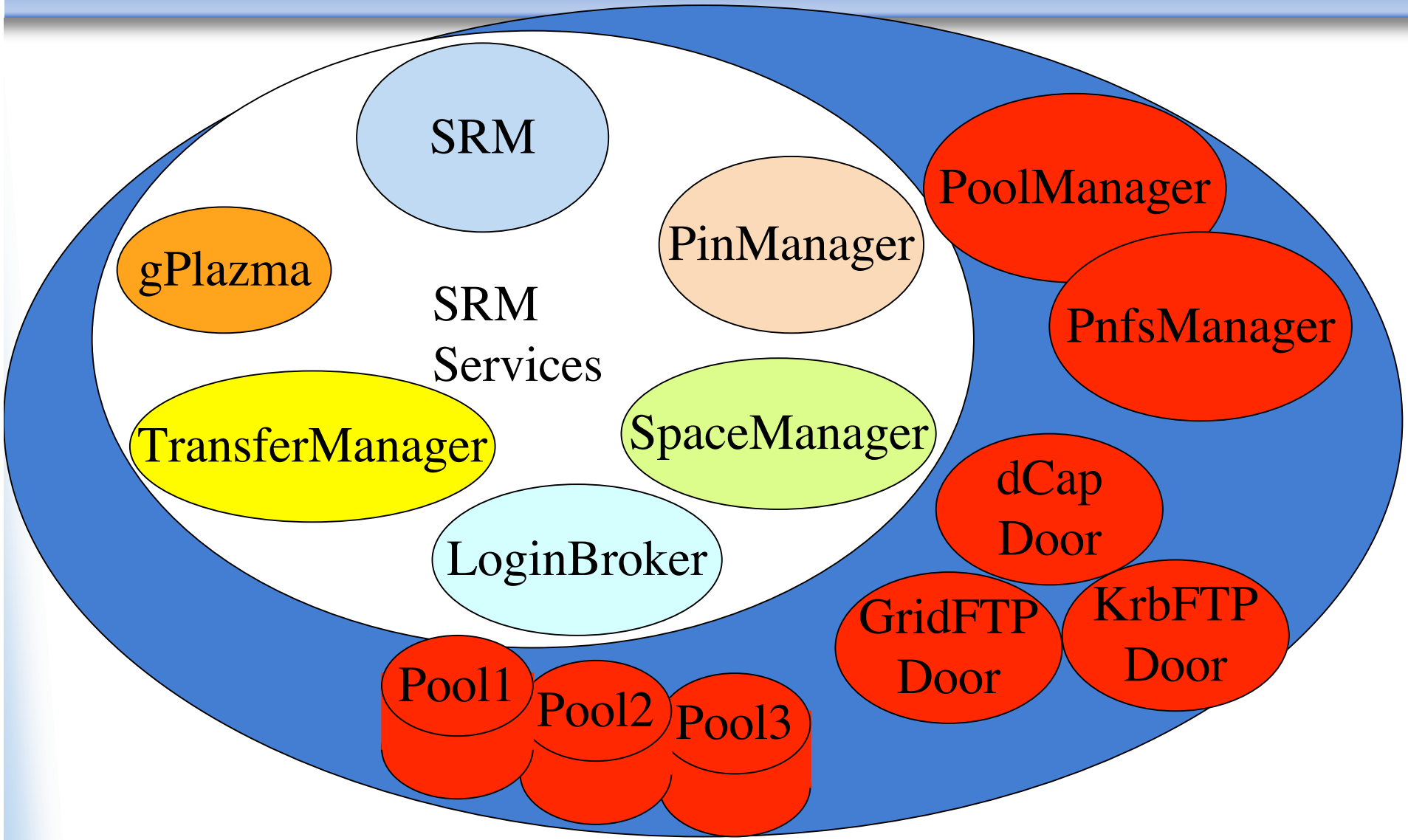
dCache SRM v2.2 history

- Prototype of SRM 2.0 interface demonstrated at SC 2003
- Work on dCache SRM 2.x since late 2004
- LHC experiments input led to SRM 2.2 definition
- In May 2006 WLCG chose a subset of SRM v2.2 which became a dCache project target
- Beta of dCache 1.8 with SRM 2.2 released in April 2007
- October 2007 – Production release

SRM dCache Services

- LoginBroker – a service for the discovery of all dCache Doors (a transfer protocol incarnation deployed on a given host:port) and their loads
- PinManager – a service for staging and pinning files (Control of online state)
- Space Manager – a Service that supports advanced Space Reservation and management
- Pool Repository and Namespace are modified to better support “pin in Cache” operation and “Online” file parameters
- Transfer Managers supervise “srmCopy” storage to storage transfers
- gPlazma Grid Access control service provides support for VO Based Authorization

SRM Services in DCache



LoginBroker

- SRM supports protocol negotiation, dynamic TURL allocation and Load Balancing
- dCache LoginBroker
  - Knowledge of all protocol access points a.k.a. dCache Doors
  - Advertises door instant load and maximum capacity
- dCache SRM uses this to discover supported protocols and calculate TURL of the least loaded access point (host:port)
- Support of new protocols does not require dCache SRM modifications

PinManager

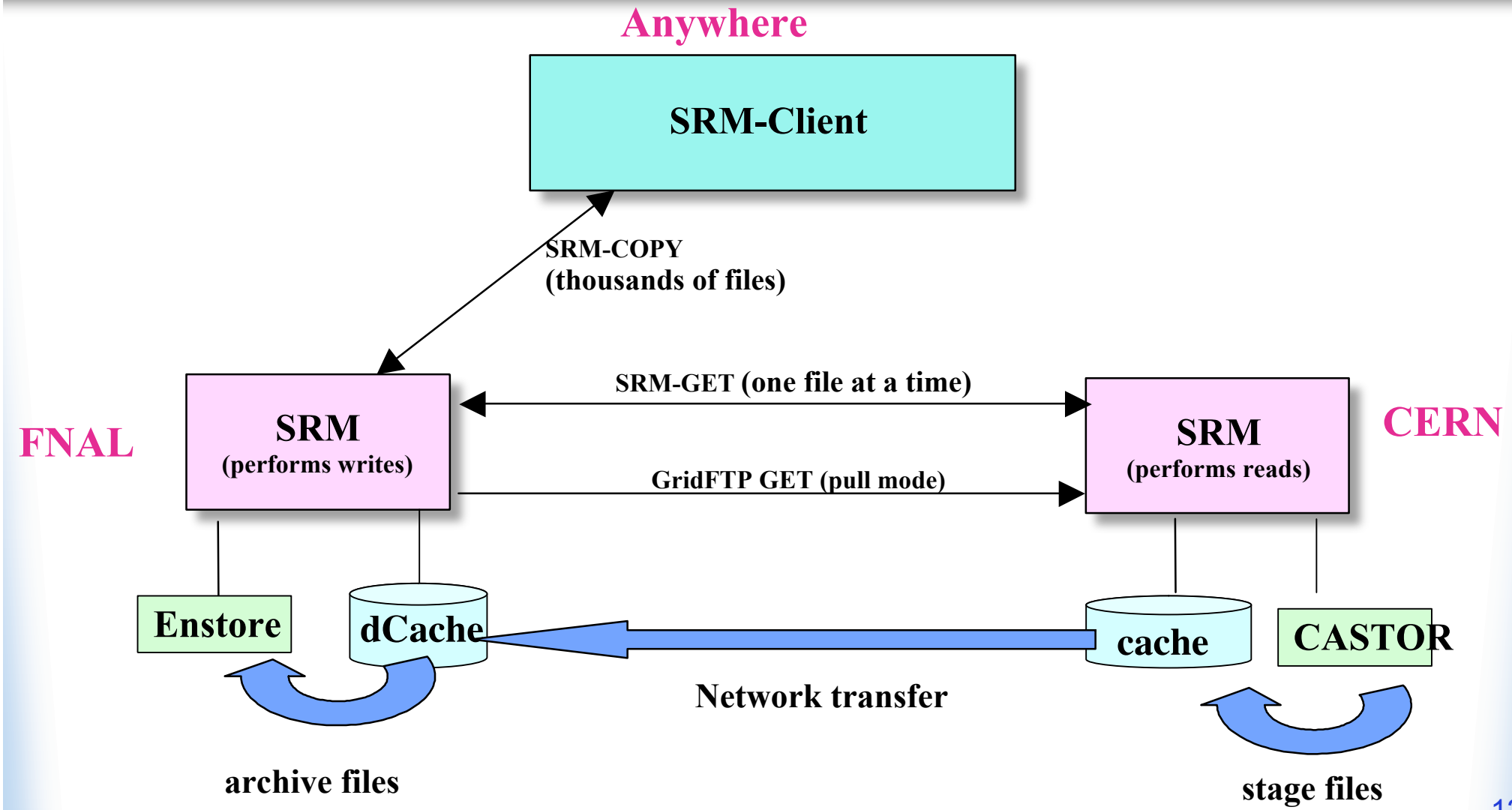
- SRM V1.1/2.2 implicit Pinning and SRM V2.2 BringOnline functionality requires advanced HSM cache management capabilities
- dCache PinManager
  - Stage and Pin functionality
  - Supports Pin lifetimes
  - Multiple logical pins mapped to a single dCache pin
  - Interface between SRM and dCache Pools and PoolManager

dCache support of Space Management

- SRM V2.2 Space Management functionality
  - Space Reservation
    - Size
    - Lifetime
    - AccessLatency (Online, Nearline) – Speed of access to the data
    - RetentionPolicy (Replica, Output, Custodial)*– quality of retention service
  - dCache had to
    - Update PoolManager pool selection mechanism
    - New Pool Repository code
    - SpaceManager module
      - dCache introduced a new way to partition total space according to their support of RetentionPolicy, AccessLatency and VO Groups/Roles.
      - Support for a streaming to HSM model

* WLCG interpretation: Replica – Disk, Custodial –Tape, Online Output – not used.

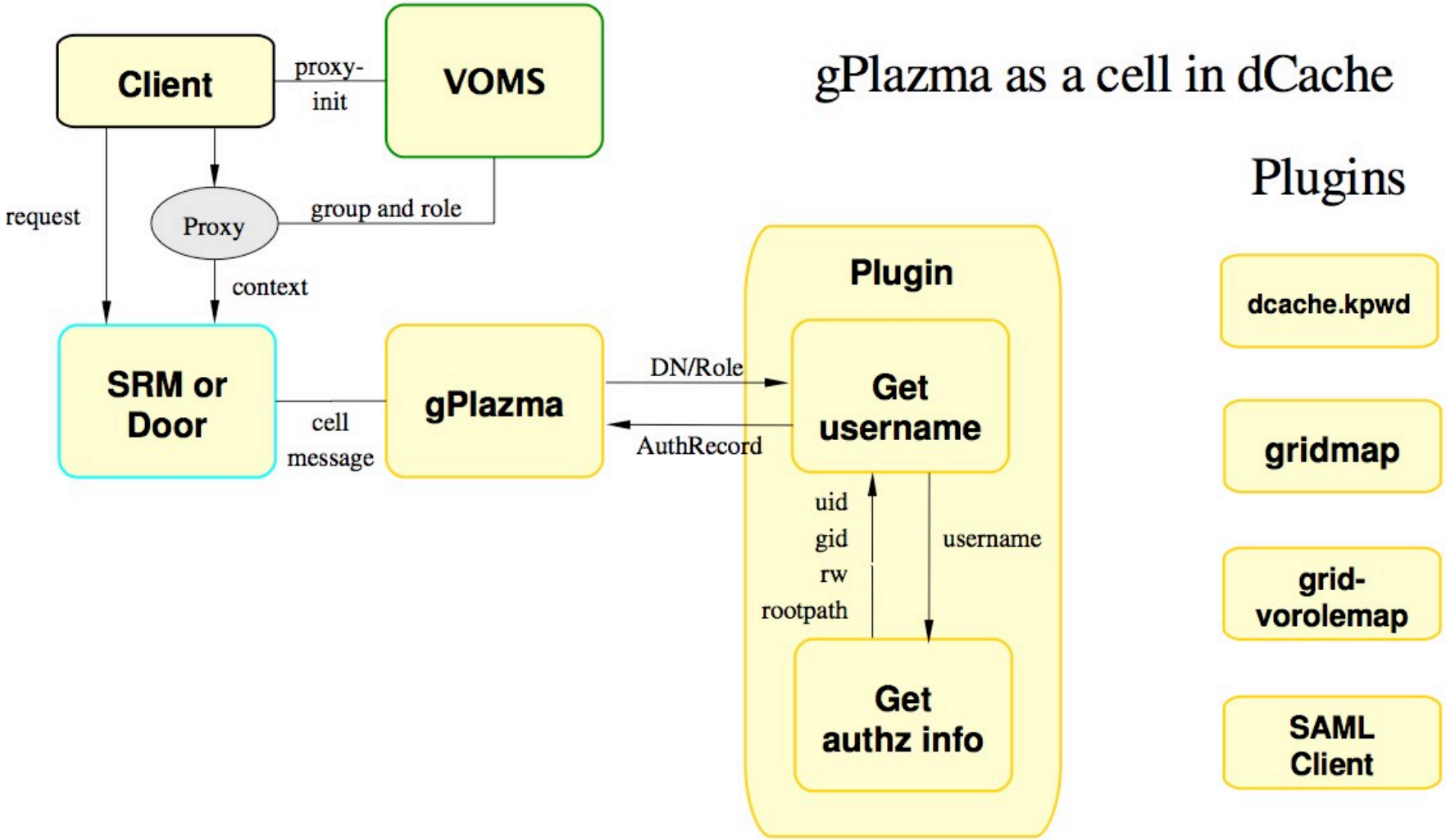
Robust Replication



Transfer Managers

- srmCopy function allows for a direct Storage to Storage transfer
- SRM Performs protocol and TURL negotiation with remote system using SrmPrepareToGet or SrmPrepareToPut functions
- dCache Transfer Managers manage individual transfers
  - Transfers are scheduled on dCache pools providing high scalability and transfer rates
  - Credential Management and delegation
  - FTP, Gridftp and http protocol support

gPlazma


Client
proxy-
init
VOMS
request
Proxy
group and role
context
SRM or
Door
cell
message
gPlazma
DN/Role
AuthRecord
Plugin
Get
username
uid
gid
rw
rootpath
username
Get
authz info
gPlazma as a cell in dCache
Plugins
dcache.kpwd
gridmap
grid-
vorolemap
SAML
Client

Access Control Mechanisms

- dCache Authorization (gPlazma)
  - Pluggable authorization service
  - Supports VO Certificate proxies
  - Multiple VO Memberships
- PNFS Namespace Service
  - Files are owned by a particular User and Group.
  - No ACL Support
- Chimera Namespace (currently in Beta testing)
  - Full ACL Support by Fall 2007
- SRM permission management functions
  - need both VO Authorization System and ACL capable Namespace Service
  - Full support of SRM Permission Management will follow

SRM 2.2 Status (1)

Basic WLCG MoU functionality

Missing 0 out of 25

WLCG MoU functionality due end of 2007

Missing 2 out of 4

Non MoU functionality

Missing 6 out of 12

Project internal stress tests show greatly improved performance. Need independent confirmation

SRM 2.2 Status (2)

- 1.8.0_14 released on Aug 24, 07
  - Resolved most URGENT issues
    - SrmMkdir, SrmRmDir, SrmRm Error Codes
    - Resolved Space Reservation VO base Access control and Performance issues
    - Support for Infinite Space Reservations, admin commands for Space Reservations in LinkGroups
    - Stability under stress tests

SRM Versions support in dCache 1.8

- **dCache 1.8.0** and higher will support **SRM 1.1 and SRM 2.2** at the same time on the same TCP Port.
- Both SRM protocol versions will run in the same dCache instance, using just one file system instance. (pnfs)
- Both SRM versions will have access to the **same file name space.**
- Files written with 1.1 can be accessed via 2.2 and vice versa.

Plans

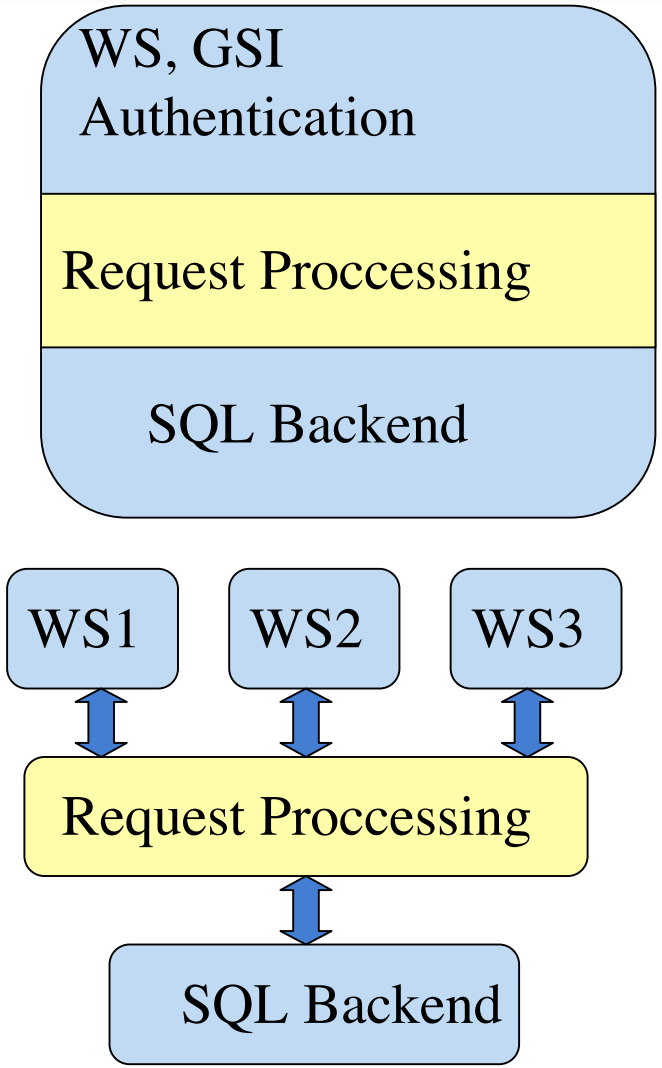
- Stability, WLCG Functionality
- SRM horizontal scaling
- Automatic File Removal after Space Expiration
- Permission Management
- Production quality srmcp client tool suite

Stability and Completion of WLCG Functionality

- First goal – stability and production quality of service provided by dCache 1.8 with SRM 1.1 and SRM 2.2
- SRM Scalability concerns will be addressed
  - Comprehensive stress testing
  - BNL – Horizontal Scalability
  - NDGF – GSI SSL optimization
- Implementation of srmChangeSpaceForFiles – early 2008

SRM Horizontal Scaling (BNL Contribution)

- SRM Interface dCache
  - WEB Service deployed in Tomcat/Axis
  - SQL database for Persistent State Storage
  - Monolith module
  - GSI Authentication – 90% CPU load
  - Does not scale to multi-nodes
- Work just began
  - Decuple Web Service interface from Business Logic
  - Allow multiple WS endpoints for a single system
  - This will enable usage of DNS Load Balancing



Automatic File Removal after Space Expiration (OSG Contribution)

- Open Science Grid (OSG) storage is open to opportunistic use by multiple experiments and organizations
- Requires ability to Guarantee that upon the expiration of the lease on disk space, it will be automatically cleaned up
  - Files in the expired Space will be removed (garbage collected) via administrative tools (or automatically)
- OSG Contribution will help add support for volatile files with managed lifetimes
- SRM SpaceManager will be used for enabling this functionality

SRM Permission Management

- need both VO Authorization System (gPlazma, done) and ACL capable Namespace Service (Chimera, Fall 07)
- Full support of SRM Permission Management will follow

SRMCP Client tool suite

- SRMCP tool suite includes
  - Srm V1.1 clients – production quality
  - Srm V2.2 clients – requires some development, documentation
- dCache 1.7 rpm clients **will not** work with dCache 1.8 SRM 2.2 interface (will work with 1.8 SRM V1.1)
- dCache 1.8 rpm client will work with dCache 1.7 SRM V1.1 interface and dCache 1.8 SRM V1.1 and V2.2 interfaces

Conclusion

- dCache team made good progress towards delivering required functionality
- Achieving production quality of the dCache 1.8 and SRM 2.2 is a highest priority
- New members (BNL, NDGF, OSG, …) strengthen collaboration, expected to contribute to the long success

References

- dCache www.dcache.org
- dCache SRM http://srm.fnal.gov
- SRM Working Group http://sdm.lbl.gov/srm-wg/
- SRM V2.2 spec http://sdm.lbl.gov/srm-wg/doc/SRM.v2.2.html